

Request to Archive
With The National Centers for Environmental Information
For Zonal-mean data set of global atmospheric reanalyses on pressure levels
Provided by UC Los Angeles

2017-04-04

This information will be used by NCEI to conduct an appraisal and make a decision on the request.

1. Who is the primary point of contact for this request?

Patrick Martineau
UC Los Angeles
Postdoctoral Scholar
patrick.martineau2@mail.mcgill.ca

2. Name the organization or group responsible for creating the dataset.

Myself, contributor to S-RIP project (SPARC Reanalysis Intercomparison Project)

3. Provide an overview summarizing the scope of data you want to archive. Describe the outputs, data variables, including their measurement resolution and coverage.

The dataset is a collection of zonal-mean atmospheric diagnostics computed from reanalysis datasets on pressure levels. Primary variables include temperature, geopotential height and the three-dimensional wind field. Advanced diagnostics provide zonal covariance terms that can be used to compute, for instance, eddy kinetic energy and eddy fluxes. Terms from the primitive zonal-mean momentum equation and the transformed Eulerian momentum equations are also provided. All diagnostics are provided in function of latitude (from pole to pole) and pressure (from surface to each reanalysis' uppermost level) from 1958 to present, depending each reanalysis' availability. The dataset is provided for two distinct grid type. The first provides all diagnostics on the grid of each reanalysis' raw data. The second provides diagnostics on a common grid of 2.5 by 2.5 degrees latitude and longitude after linearly interpolating the raw data. This dataset is produced to facilitate the comparison of reanalysis datasets for the collaborators of the SPARC-Reanalysis Intercomparison Project (S-RIP) project and the atmospheric science community at large. The dataset is substantially smaller in size compared to the full three dimensional reanalysis fields and uses unified numerical methods. The dataset includes all global reanalyses available at the time of its development and will be extended to new datasets in the future. 1958-2012 are the official dates for which the dataset is currently available which means that all reanalysis products that extend back to 1958 are provided back to 1958 and all reanalysis that are provided after 2012 are included up to 2012. Providing the dataset beyond 2012 is an ongoing process and availability for all reanalyses is not guaranteed.

4. What is the time period covered by the dataset? (YYYY-MM-DD, YYYY-MM or YYYY)

From 1958
Ongoing as continuous updates to the data record

5. Edition or version number(s) of the dataset:

I am now providing version-2 to the S-RIP community and have discontinued support of version-1.

6. Approximate date when the dataset was or will be released to the public:

2017-09-01

7. Who are the expected users of the archived data? How will the archived data be used?

S-RIP community and the climate dynamics community at large. Although the dataset was developed for the S-RIP project, some have shown interest using the dataset outside of the S-SRIP project. The diagnostics provided are those commonly used by the climate dynamics community. A reference paper and a stable server to archive the dataset as well as a DOI will play an important role in sharing and spreading the word about the dataset.

8. Has the dataset undergone user evaluation and/or an independent review process? Did NCEI participate in design reviews?

Several contributors to the S-RIP project are using the dataset in their own contributions which helped identify and correct minor mistakes at the beginning of the project. An independent review will be carried when submitting a manuscript describing the dataset to ESSD journal, but this requires that the data be uploaded to a stable server and be assigned a DOI before submission of the manuscript. NCEI did not participate in design reviews.

9. Describe the dataset's relationship to other archived datasets, such as earlier versions or related source data. If this is a new version, how does it improve upon the previous version(s)?

The dataset provides zonal-mean diagnostics of reanalysis datasets. The benefits of this dataset include 1) the ability of quickly comparing reanalysis data without the need to download the full 3D fields (1.5 Tb instead of 25+ Tb) and 2) the provision of advanced diagnostics often used by the climate dynamics community are already computed, saving time to the users.

10. List the input datasets and ancillary information used to produce the data.

The zonal-mean data set on pressure levels includes most major reanalysis products. Three reanalyses are provided by the European Centre for Medium-Range Weather Forecasts (ECMWF): 40-year Reanalysis (ERA-40), ECMWF Interim Reanalysis (ERA-interim) and ECMWF Atmospheric Reanalysis of the 20th Century (ERA-20C).

Reanalyses from National Oceanic and Atmospheric Administration (NOAA) include the NOAA's National Centers for Environmental Prediction/National Center for Atmospheric Research reanalysis (NCEP-NCAR) and its successor the NCEP-DOE Atmospheric Model Intercomparison Project (AMIP) reanalysis (NCEP-DOE), the NCEP Climate Forecast System Reanalysis (NCEP-CFSR) and NOAA's 20th Century Reanalysis (20CR) version 2 (v2) and 2c (v2c). Note that the NCEP-CFSR data after January 2011 have been produced with a slightly different system CFSv2.

Four reanalyses come from the Japanese Meteorological Agency (JMA): the Japanese 25-year Reanalysis (JRA-25), the Japanese 55-year Reanalysis (JRA-55) and two variants (JRA-55C, JRA-55AMIP). The first one, JRA-55C, assimilates only conventional surface and upper air observations while the second JRA-55AMIP is an AMIP-type simulation.

Finally, two reanalyses are from the National Aeronautics and Space Administration (NASA): the Modern-Era Retrospective-analysis for Research and Applications version one (MERRA) and two (MERRA-2).

11. List web pages and other links that provide information on the data.

<http://p-martineau.com/s-rip/>

12. List the kinds of documents, metadata and code that are available for archiving. For example, data format specifications, user guides, algorithm documentation, metadata compliant with a standard such as ISO 19115, source code, platform/instrument metadata, data/process flow diagrams, etc.

1. The dataset is provided as NetCDF files that are CF compliant whenever possible. A user guide is available for the dataset and a manuscript describing the dataset will be submitted to ESSD (both can be supplied on demand).

13. Indicate the data file format(s).

1. netCDF-4

14. Are the data files compressed?

No

15. Provide details on how the files are named and how they are organized (e.g., file_name_pattern_YYYYMM.tar in monthly aggregations).

Example of filename:

ZMD_4xdaily_yyyy_mm.nc in monthly aggregations.

16. Explain how to access sample data files and/or a file listing for previewing. If it is not available now, when will it be available?

The dataset is available at http://meteo.mcgill.ca/srip/common_grid_v2/

17. What is the total data volume to be submitted?

Historic Data: all historic data or data submitted as a completed collection.

Total Data Volume: 1.5TB

Number of Data Files: 100000

18. Are later updates, revisions or replacement files anticipated? If so, explain the conditions for submitting these additional data to the archive.

The dataset will be updated as reanalyses datasets are extended to include latest years.

19. Describe the server that will connect to the ingest server at NCEI for submitting the data.

Physical Location: Los Angeles, California

System Name: Athens

System Owner: Prof. Gang Chen, my supervisor at UCLA

Additional Information:

20. What are the possible methods for submitting the data to NCEI? Select all that apply.

SSH, RSYNC

21. Identify how you would like NCEI to distribute the data. Web access support depends on the resources available for the dataset.

1. Direct download links

22. Will there be any distribution, usage, or other restrictions that apply to the data in the archive?

No known constraints apply to the data.

23. Discuss the rationale for archiving the dataset and the anticipated benefits. Mention any risks associated with not archiving the dataset at NCEI.

Contributors to the S-RIP project have voiced strong interest to see a manuscript published about the dataset. To this end, the data needs to be uploaded to a safe and stable server. At the moment, the data is published on a server at McGill university, and produced on a computer at UCLA. It is uncertain whether the data will remain at these storage facilities as they are not official and are only provided to me temporarily. Since I may still change affiliation in the future as my appointment is temporary, there is a risk that I may not be able to publicly distribute the dataset if I am

not able to secure enough computer storage space and a stable server in the future. This could potentially delay the progress of several contributors to S-RIP.

24. Are the data archived at another facility or are there plans to do so? Please explain.

The data is available on my research computer in the format of a web file listing:

http://meteo.mcgill.ca/srip/common_grid_v2/

http://meteo.mcgill.ca/srip/original_grid_v2/

but the stability of this server is not guaranteed and the computer is aging.

In case NCEI accepts to host the dataset, a copy of the dataset will be kept only on the computer used to produce the dataset.

25. Is there an existing agreement or requirement driving this request to archive? Have you already contacted someone at NCEI?

No

26. Do you have a data management plan for your data?

No

27. Have funds been allocated to archive the data at NCEI?

No

28. Identify the affiliated research project, its sponsor, and any project/grant ID as applicable.

The SPARC Reanalysis Intercomparison Project (S-RIP)

<https://s-rip.ees.hokudai.ac.jp/index.html>

29. Is there a desired deadline for NCEI to archive and provide access to the data?

Archive by: 2017-09-01

Accessible by: 2017-09-30

30. Add any other pertinent information for this request.

None